Nonlinear programming with cumulatively bounded variables

Lazaros P. Mavrides (*)

ABSTRACT

The problem [maximize $f(x)$, subject to $x_1 + \ldots + x_j \geqslant b_j$ for $j = 1,\ldots, N$] is solved by a feasible direction method that takes advantage of its special structure. A direction vector that approximates the vector of Lagrange multipliers is used. In the one-dimensional subproblem the direction vector is bent every time a constraint becomes active. Convergence to a K–T point is proven. McCormick has used a similar method for the problem [maximize $f(x)$, subject to $x \geqslant 0$], with the gradient as direction vector. A computationally implementable algorithm is given, with a finite stepsize procedure and a finite stopping rule. Observations from numerous applications to a recurring banking problem are discussed. Related techniques might be useful in other situations.

1. INTRODUCTION

We consider the problem

$$\text{maximize } f(x), \tag{1}$$

$$\text{subject to } x_1 + \ldots + x_j \geqslant b_j \quad \text{for all } j, \tag{2}$$

where $f : R^N \to R$ is continuously differentiable on the feasible region.
It can be seen that the K–T conditions for this problem can be expressed as follows : If x^* is optimal, then there exists a multiplier vector $\lambda^* \geqslant 0$ such that for $j = 1,\ldots, N$

1) $x_1^* + \ldots + x_j^* \geqslant b_j$,

2) $\lambda_j^* (x_1^* + \ldots + x_j^* - b_j) = 0$,

3) $\partial f(x^*)/\partial x_j + \lambda_j^* + \ldots + \lambda_N^* = 0$.

Note that K–T condition 3 holds if and only if

$$\lambda_j^* = \partial f(x^*)/\partial x_{j+1} - \partial f(x^*)/\partial x_j \quad \text{for all } j, \tag{3}$$

with

$$\partial f(x)/\partial x_{N+1} \equiv 0.$$

The method to be discussed in the next section begins in each iteration with the unique vector of multipliers (not necessarily nonnegative) that together with the point of iteration satisfies K–T condition 3 (determined by (3)). In its one-dimensional subproblem the method proceeds by bending the direction vector every time a constraint becomes active. The bending is done in such a way that movement ceases along the components that violate feasibility, but continues along the other components.
McCormick [8] has proposed a similar method for the problem [maximize $f(x)$, subject to $x \geqslant 0$], using the gradient of the objective function at the point of iteration as the initial direction vector. It can be shown that McCormick's method cannot be applied to the problem defined by (1) and (2), due to the possibility of zigzagging (alias jamming; for a discussion of this problem see Zangwill [9], p. 279).

2. THE METHOD

A theoretically convergent version of the method will be discussed first. A finite step-size procedure and a finite stopping rule will be given subsequently.

Algorithm

Step 0
Begin with some feasible point $x^{(0)}$. Set $n = 0$.

Step 1
Compute the initial direction vector $\lambda^{(n)}$, where

$$\lambda_j^{(n)} = \partial f(x^{(n)})/\partial x_{j+1} - \partial f(x^{(n)})/\partial x_j \tag{4}$$

for all j,

with

$$\partial f(x)/\partial x_{N+1} \equiv 0.$$

Step 2
(one-dimensional subproblem). For $\theta \in R$, let

$$x_j(\theta) = \max \{b_j - \sum_{i=1}^{j-1} x_i(\theta),\ x_j - \theta\lambda_j\} \quad \text{for all } j. \tag{5}$$

(*) L. P. Mavrides, Morgan Guaranty Trust Company, 23 Wall Street, New York, N.Y. 10015, U.S.A.

Then determine $\theta^{(n)}$, an optimal solution of the problem

$$\text{maximize } f[x^{(n)}(\theta)], \tag{6}$$

$$\text{subject to } 0 \leqslant \theta \leqslant \bar{\theta}, \tag{7}$$

where $\bar{\theta} > 0$ is a user-specified constant.

Step 3.

Set $x^{(n+1)} = x^{(n)}[\theta^{(n)}]$. Replace n by n+1 and go to Step 1.

This algorithm was developed by relating an approach in Cazalet [1] with the algorithm by McCormick [8].

3. CONVERGENCE

Lemma 1

For any feasible point x, $df[x(0)]/d\theta^+ \geqslant 0$ and, if x is not a K–T point, then $df[x(0)]/d\theta^+ > 0$.

Proof

Clearly, with $\theta \geqslant 0$,

$$dx_j(\theta)/d\theta^+ = \begin{cases} -\lambda_j \\ \quad \text{if } x_j - \theta\lambda_j > b_j - \sum_{i=1}^{j-1} x_i(\theta), \\ -\sum_{i=1}^{j-1} dx_i(\theta)/d\theta^+ \\ \quad \text{if } x_j - \theta\lambda_j < b_j - \sum_{i=1}^{j-1} x_i(\theta), \\ \max\{-\sum_{i=1}^{j-1} dx_i(\theta)/d\theta^+, -\lambda_j\} \\ \quad \text{if } x_j - \theta\lambda_j = b_j - \sum_{i=1}^{j-1} x_i(\theta), \end{cases} \tag{8}$$

for $j = 1,\ldots,N$. Consequently, if x is a feasible point, then

$$dx_j(0)/d\theta^+ = \begin{cases} -\lambda_j \\ \quad \text{if } x_j > b_j - \sum_{i=1}^{j-1} x_i, \\ \max\{-\sum_{i=1}^{j-1} dx_i(0)/d\theta^+, -\lambda_j\} \\ \quad \text{if } x_j = b_j - \sum_{i=1}^{j-1} x_i, \end{cases} \tag{9}$$

from which it follows that, either $dx_j(0)/d\theta^+ = -\lambda_j$ for any j, or there exist $j_1,\ldots,j_n$ with $1 \leqslant j_1 < \ldots < j_n \leqslant N$ such that for $k = 1,\ldots,n$

$$x_{j_k} = b_{j_k} - \sum_{i=1}^{j_k-1} x_i \text{ and } -\sum_{i=1}^{j_k-1} dx_i(0)/d\theta^+ > -\lambda_{j_k} \tag{10}$$

in which case

$$dx_j(0)/d\theta^+ = \begin{cases} -\sum_{i=1}^{j-1} dx_i(0)/d\theta^+ > -\lambda_j \\ \quad \text{for } j = j_1,\ldots,j_n, \\ -\lambda_j \quad \text{otherwise.} \end{cases} \tag{11}$$

Clearly, for $k = 1,\ldots,n$,

$$dx_{j_k}(0)/d\theta^+ \begin{cases} = -\sum_{j=1}^{j_{k-1}-1} dx_j(0)/d\theta^+ \\ \quad -dx_{j_{k-1}}(0)/d\theta^+ - \sum_{j=j_{k-1}+1}^{j_k-1} dx_j(0)/d\theta^+, \\ = \sum_{j=j_{k-1}+1}^{j_k-1} dx_j(0)/d\theta^+, \end{cases} \tag{12}$$

which implies that

$$dx_j(0)/d\theta^+ = \begin{cases} -\lambda_j & \text{for } j \neq j_1,\ldots,j_n, \\ \sum_{i=j_{k-1}+1}^{j_k-1} \lambda_i & \text{for } j = j_k \text{ and } k=1,\ldots,n. \end{cases} \tag{13}$$

Letting λ be as in (4) for any feasible point x and substituting this in (13), we have, with $j_0 \equiv 0$,

$$dx_j(0)/d\theta^+ = \begin{cases} \partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1} \\ \quad \text{for } j \neq j_1,\ldots,j_n, \\ \partial f(x)/\partial x_{j_k} - \partial f(x)/\partial x_{j_{k-1}+1} \\ \quad \text{for } j = j_k \text{ and } k=1,\ldots,n. \end{cases} \tag{14}$$

Clearly, since $x(0) = x$,

$$df[x(0)]/d\theta^+ = \sum_{j=1}^{N} \partial f(x)/\partial x_j \, \partial x_j(0)/d\theta^+ \cdot \tag{15}$$

Consequently,

$$df[x(0)]/d\theta^+ = \sum_{\substack{j=1 \\ j\neq j_1,\ldots,j_n}}^{N} \partial f(x)/\partial x_j[\partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1}]$$
$$+ \sum_{k=1}^{n} \partial f(x)/\partial x_{j_k}[\partial f(x)/\partial x_{j_k} - \partial f(x)/\partial x_{j_{k-1}+1}] \tag{16}$$

Note that for all k and ρ, with $1 \leqslant k < \rho \leqslant N$,

$$\sum_{j=k}^{\rho} \partial f(x)/\partial x_j[\partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1}]$$

$$= \sum_{j=k}^{\rho} 1/2 \;\{[\partial f(x)/\partial x_j]^2 + [\partial f(x)/\partial x_{j+1}]^2$$

$$- 2\, \partial f(x)/\partial x_j\, \partial f(x)/\partial x_{j+1}\}$$

$$+ 1/2\; \{[\partial f(x)/\partial x_k]^2 - [\partial f(x)/\partial x_{\rho+1}]^2\}$$

$$= 1/2\; \{[\partial f(x)/\partial x_k]^2 - [\partial f(x)/\partial x_{\rho+1}]^2$$

$$+ \sum_{j=k}^{\rho} [\partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1}]^2\}\;, \tag{17}$$

where $\partial f(x)/\partial x_{N+1} \equiv 0$. Hence,

$$df[x(0)]/d\theta^+ = 1/2\; \{[\partial f(x)/\partial x_1]^2$$

$$- [\partial f(x)/\partial x_{j_1}]^2 + \sum_{j=1}^{j_1-1} [\partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1}]^2\}$$

$$+ \partial f(x)/\partial x_{j_1}\, [\partial f(x)/\partial x_{j_1} - \partial f(x)/\partial x_1]$$

$$+ \;\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+ 1/2\; \{[\partial f(x)/\partial x_{j_{n-1}+1}]^2 - [\partial f(x)/\partial x_{j_n}]^2$$

$$+ \sum_{j=j_{n-1}+1}^{j_n-1} [\partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1}]^2\}$$

$$+ \partial f(x)/\partial x_{j_n}\, [\partial f(x)/\partial x_{j_n} - \partial f(x)/\partial x_{j_{n-1}+1}]$$

$$+ 1/2\; \{[\partial f(x)/\partial x_{j_n+1}]^2 - [\partial f(x)/\partial x_{j_{n+1}}]^2$$

$$+ \sum_{j=j_n+1}^{j_{n+1}-1} [\partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1}]^2\}$$

$$+ \partial f(x)/\partial x_{j_{n+1}} \cdot \partial f(x)/\partial x_{j_{n+1}}\;, \tag{18}$$

with

$$j_{n+1} \equiv \begin{cases} N & \text{if } j_n < N \\ N+1 & \text{if } j_n = N\,. \end{cases}$$

Upon rearrangement and cancellation of terms in (18),

$$df[x(0)]/d\theta^+ \begin{cases} = 1/2\; \{\sum_{k=1}^{n} [\partial f(x)/\partial x_{j_k} - \partial f(x)/\partial x_{j_{k-1}+1}]^2 \\ + [\partial f(x)/\partial x_{j_n+1}]^2 + [\partial f(x)/\partial x_{j_{n+1}}]^2 \\ + \sum_{\substack{j=1 \\ j \neq j_1,\ldots,j_n}}^{N-1} [\partial f(x)/\partial x_j - \partial f(x)/\partial x_{j+1}]^2\} \\ \geqslant 0. \end{cases} \tag{19}$$

Clearly, if for $j = 1,\ldots, N$, either

$$\partial f(x)/\partial x_j = \partial f(x)/\partial x_{j+1} \text{ and } x_j > b_j - \sum_{i=1}^{j-1} x_i\,, \tag{20}$$

or

$$\partial f(x)/\partial x_j \leqslant \partial f(x)/\partial x_{j+1} \text{ and } x_j = b_j - \sum_{i=1}^{j-1} x_i, \tag{21}$$

then (x, λ), with λ computed as in (4), will satisfy the K–T conditions. Therefore, if x is a K–T point, then, either there exists a j such that

$$\partial f(x)/\partial x_j \neq \partial f(x)/\partial x_{j+1} \text{ and } x_j > b_j - \sum_{i=1}^{j-1} x_i, \tag{22}$$

or there exists a j such that

$$\partial f(x)/\partial x_j > \partial f(x)/\partial x_{j+1} \text{ and } x_j = b_j - \sum_{i=1}^{j-1} x_i. \tag{23}$$

From (10), $x_j = b_j - \sum_{i=1}^{j-1} x_i$ for $j = j_1,\ldots,j_n$. Hence, if x is not a K–T point, then either there exists a $j \neq j_1,\ldots,j_n$ such that $\partial f(x)/\partial x_j \neq \partial f(x)/\partial x_{j+1}$, or there exists a $j \in \{j_1,\ldots, j_n\}$ such that $\partial f(x)/\partial x_j > \partial f(x)/\partial x_{j+1}$.

Suppose that x is not a K–T point and $\partial f(x)/\partial x_j = \partial f(x)/\partial x_{j+1}$ for $j \notin \{j_1,\ldots,j_n\}$. From (11) and (14),

$$dx_{j_k}(0)/d\theta^+ \begin{cases} = - \sum_{i=1}^{j_k-1} dx_i(0)/d\theta^+ \\ \quad = \partial f(x)/\partial x_{j_k} - \partial f(x)/\partial x_{j_{k-1}+1} \\ > -\lambda_{j_k} \\ \quad = \partial f(x)/\partial x_{j_k} - \partial f(x)/\partial x_{j_k+1}\,. \end{cases} \tag{24}$$

It follows that $\partial f(x)/\partial x_{j_{k-1}+1} < \partial f(x)/\partial x_{j_k+1}$ for $k = 1,\ldots, n$. Since $j \notin \{j_1,\ldots,j_n\}$ for $j = j_{k-1}+1,\ldots,j_k-1$, $\partial f(x)/\partial x_j = \partial f(x)/\partial x_{j+1}$ for these j by assumption. It follows that $\partial f(x)/\partial x_{j_k} < \partial f(x)/\partial x_{j_k+1}$ for $k = 1,\ldots, n$. Consequently, for $j = 1,\ldots, N$, either (20) or (21) holds and x must be a K–T point, thus contradicting the initial assumption.

Hence, if x is not a K–T point, then there exists a $j \notin \{j_1,\ldots,j_n\}$ such that $\partial f(x)/\partial x_j \neq \partial f(x)/\partial x_{j+1}$. In that case (19) becomes a strict inequality. (If $j = N$ is the only such j, then $j_n < N$ and $j_{n+1} = N$ and, thus, $[\partial f(x)/\partial x_{j_{n+1}}]^2 > 0$.) Q.E.D.

Lemma 2

For any given $\theta \in R$, $x(\theta)$ (as determined in (5)) and $f[x(\theta)]$ are both continuous over $x \in F^o$.

Proof

By assumption, f is continuously differentiable on F^o. Thus, $\nabla f(x)$ is continuous over $x \in F^o$, implying that λ (determined as in (4)) is continuous over $x \in F^o$. Hence, $x - \theta\lambda$ is continuous over $x \in F^o$. It follows that

$$x_1(\theta) = \max \{b_1, x_1 - \theta\lambda_1\} \tag{25}$$

is continuous over $x \in F^o$.

For induction purposes, assume that $x_j(\theta)$ is continuous over $x \in F^o$ for some $j < k$. Then $b_k - \sum_{i=1}^{k-1} x_i(\theta)$ and $x_k - \theta\lambda_k$ are both continuous over $x \in F^o$ and, thereby,

$$x_k(\theta) = \max \{b_k - \sum_{i=1}^{k-1} x_i(\theta),\ x_k - \theta\lambda_k\} \tag{26}$$

is continuous over $x \in F^o$. By induction, $x_j(\theta)$ is continuous over $x \in F^o$.

Clearly, if $x \in F^o$, then $x(\theta) \in F^o$. Therefore, $x(\theta)$ and $f(x)$ both continuous over $x \in F^o$ imply that $f[x(\theta)]$ is continuous over $x \in F^o$. Q.E.D.

Lemma 3

If $x' \in F^o$ is not a K–T point, then for any $\epsilon > 0$ there exists a $\sigma > 0$ such that

$$\max_{0 \leqslant \theta \leqslant \epsilon} f[x(\theta)] > f(x') \text{ for } x \in B(x', \sigma) \cap F^o \tag{27}$$

where $B(x', \sigma)$ denotes the open ball centered at x' with radius σ.

Proof

Suppose that the assertion is not true. Then there must exist $\epsilon > 0$ and a sequence $\{x^n\}_0^\infty \subset F^o$ such that $x^n \to x'$ and $f[x^n(\theta)] \leqslant f(x')$ for $\theta \in [0, \epsilon]$. If so, then $f[x(\theta)]$ continuous over $x \in F^o$ (by Lemma 2) implies that $f[x'(\theta)] - f(x') \leqslant 0$ and, thereby,

$$df[x'(0)]/d\theta^+ \equiv \lim_{\theta \to 0^+} \{f[x'(\theta)] - f(x')\}/\theta \leqslant 0, \tag{28}$$

a contradiction, since, by Lemma 1, if x' is not a K–T point, then $df[x'(0)]/d\theta^+ > 0$. Q.E.D.

Theorem 1

Suppose $B(x^{(0)}) \equiv \{x \in F^o \mid f(x) \geqslant f(x^{(0)})\}$ is bounded. Then $\{f(x^{(n)})\}_0^\infty$ has a limit, $\{x^{(n)}\}_0^\infty$ has at least one limit point, and every limit point of $\{x^{(n)}\}_0^\infty$ is a K–T point.

Proof

F^o is the intersection of closed half spaces and, therefore, is closed. This together with the fact that f is continuous over $x \in F^o$ (by assumption) imply that $B[x^{(0)}]$ is closed. $B[x^{(0)}]$ is also bounded (by assumption) and, therefore, it is compact.

Note that

$$f[x^{(n+1)}] \equiv f\{x^{(n)}[\theta^{(n)}]\} = \max_{0 \leqslant \theta \leqslant \bar{\theta}} f[x^{(n)}(\theta)]$$

$$\geqslant f[x^{(n)}(0)] = f[x^{(n)}]. \tag{29}$$

Hence, $\{f[x^{(n)}]\}_0^\infty$ is a bounded monotonic sequence in R and, therefore, $\lim_{n \to \infty} f[x^{(n)}] = \bar{f} < \infty$ exists.

On the other hand, $\{x^{(n)}\}_0^\infty$ is an infinite sequence in $B[x^{(0)}]$, which has been shown to be compact and, therefore, has at least one limit point.

Let x' be one limit point of $\{x^{(n)}\}_0^\infty$ and $\{x^n\}_0^\infty$ be a subsequence converging to x'. Then, by Lemma 3, if x' is not a K–T point, for x^n sufficiently close to x',

$$f[x^n(\theta^n)] = \max_{0 \leqslant \theta \leqslant \bar{\theta}} f[x^n(\theta)] > f(x'). \tag{30}$$

However, x' a limit point of $\{x^{(n)}\}_0^\infty$ implies that $f(x') = \bar{f}$, which together with the fact that $f[x^{(n+1)}] \geqslant f[x^{(n)}]$ for all n (shown earlier) imply that

$f[x^{(0)}] \leqslant f(x^{(1)}) \leqslant \ldots \leqslant f(x')$.

Thus, the assumption that x' is not a K–T point has led to a contradiction. Q.E.D.

4. A STOPPING RULE

The algorithm generates a solution sequance $\{x^{(n)}\}_0^\infty \subset F^o$. Therefore, $x^{(n)}$ satisfies K–T condition 1 for all n. Furthermore, $\lambda^{(n)}$ is generated in such a way that together with $x^{(n)}$ they satisfy K–T condition 3 for all n. Also, $x^{(n)}$ feasible for all n implies that $\sum_{i=1}^{j} x_i^{(n)} - b_j \geqslant 0$ for all n and all j. It follows that, if $\lambda_j^{(n)} \geqslant 0$ for all j and $\sum_{j=1}^{N} \lambda_j^{(n)} [\sum_{i=1}^{j} x_i^{(n)} - b_j] = 0$, then $\lambda_j^{(n)} [\sum_{i=1}^{j} x_i^{(n)} - b_j] = 0$ for all j and, thus, $x^{(n)}$ and $\lambda^{(n)}$ together also satisfy K–T condition 3 and non-negativity of the multipliers. This suggests the following stopping rule :

Stopping rule

Letting $g_j(x) \equiv \sum_{i=1}^{j} x_i - b_j$ for all j, stop if $\lambda_j^{(n)} \geqslant -\delta$ for all j and

$$\sum_{j=1}^{N} \left| \lambda_j^{(n)} \right| g_j[x^{(n)}] \leqslant \epsilon \tag{32}$$

for some user-specified constants $\delta > 0$ and $\epsilon > 0$.

Theorem 2

The algorithm with the stopping rule would terminate in a finite number of iterations at a point arbitrarily close to a K–T point.

Proof

By assumption, f is continuously differentiable on F^o. Thus, $\nabla f(x)$ is continuous over $x \in F^o$, implying that λ (determined as in (4)) is continuous over $x \in F^o$. Clearly, $g_j(x)$ is also continuous over $x \in F^o$ for all j.
Let x^* be a limit point of $\{x^{(n)}\}_0^\infty$ (it exists, by theorem 1) and $\{x^n\}_0^\infty$ be a subsequence converging to x^*. Then x^* is a K–T point (by theorem 1).
λ continuous over $x \in F^o$ implies that $\lim_{n \to \infty} \lambda^n = \lambda^* \geqslant 0$, since x^* is a K–T point and, from (3), λ^* is the unique multiplier vector that together with x^* satisfies K–T condition 3. Hence, there exists an integer N_1 such that $\lambda_j^n \geqslant -\delta$ for all j and $n \geqslant N_1$.
On the other hand, λ_j and g_j both continuous over $x \in F^o$ for any j imply that

$$\lim_{n \to \infty} \sum_{j=1}^{N} \left| \lambda_j^n \right| g_j(x^n) = \sum_{j=1}^{N} \lambda_j^* \; g_j(x^*) = 0 \qquad (33)$$

(since x^* a K–T point implies that $\lambda_j^* g_j(x^*) = 0$ for any j, by K–T condition 2). Hence, there exists an integer N_2 such that

$$\sum_{j=1}^{N} \left| \lambda_j^n \right| g_j(x^n) \leqslant \epsilon \text{ for } n \geqslant N_2. \qquad (34)$$

Clearly, both conditions in the stopping rule are satisfied for any $n \geqslant M = \max\{N_1, N_2\}$.
If $\delta = \epsilon = 0$, then the algorithm will terminate only if the K–T point x^* has been attained, generally not in a finite number of iterations. As $\delta \to 0+$ and $\epsilon \to 0+$, $M \to \infty$ and $x^M \to x^*$. Q.E.D.

5. A FINITE STEP-SIZE PROCEDURE

The one-dimensional subproblem of the algorithm (step 2) requires an exact solution. We have thus far assumed that such a solution can be obtained somehow. Such a solution could be approximated by a one-dimensional search method such as the Method of Golden Sections or the Method of Interval Bisection, assuming unimodality. However, these methods do not generally converge finitely. We propose replacing step 2 by

Step 2′

Determine an acceptable step size $\theta^{(n)}$ via the following procedure :

A. If $df[x^{(n)}(0)]/d\theta^+ = 0$, set $\theta^{(n)} = 0$ and go to step 3. Otherwise specify a constant $b \in (0,0.5)$ and go to B.

B. Specify some $\theta > 0$. If
$$\theta b\, df[x^{(n)}(0)]/d\theta^+ \leqslant f[x^{(n)}(\theta)] - f[x^{(n)}] \leqslant \theta\,(1-b)\,df[x^{(n)}(0)]/d\theta^+ ,$$
set $\theta^{(n)} = \theta$ and go to step 3. If the left inequality is violated, set $\theta_L = 0$ and $\theta_H = \theta$ and go to C. Otherwise replace θ_L by θ and θ by 2θ and repeat B.

C. Let $\theta = (\theta_L + \theta_H)/2$. If the right inequality in B is violated, replace θ_L by θ and repeat C. If the left inequality is violated, replace θ_H by θ and repeat C. Otherwise set $\theta^{(n)} = \theta$ and go to step 3.

The step-size procedure in step 2′ seeks a point θ such that the double inequality in part B is satisfied. Whereas an exact solution is needed in step 2, any point from certain real line intervals would be acceptable for step 2′. Similar modifications have been applied by Goldstein [4] for unconstrained optimization, Daniel [2] for certain gradient-like feasible direction methods, Mavrides [6] for the method by McCormick [8] for nonnegativity constraints, and Mavrides [7] for the Frank-Wolfe Method [3].
The step-size procedure in step 2′ is finite (see [6]). Furthermore, it can be shown that the algorithm of this paper with step 2 replaced by step 2′ is convergent to a K–T point, provided that the following nondegeneracy assumption holds :

Assumption

Let x^* be a limit point of the sequence of iteration points $\{x^{(n)}\}_0^\infty$ and $\{x^n\}_0^\infty$ be a subsequence converging to x^*. Suppose that a particular constraint is active at x^*. Then there exists an integer $N < \infty$ such that this constraint is active at every x^n for $n > N$.
The above assumption would exclude only situations in which a constraint is inactive at every iteration point but active at a limit point. The proof is similar to that in [6] (step-size procedure for McCormick's algorithm for the NLP problem with nonnegativity constraints). The only difference is in that the above assumption is not required in proving convergence in [6]. However, we have only shown that this assumption is sufficient (as opposed to necessary) to prove convergence with step 2′. We have not been able to conduct a counterexample and, therefore, it might still be possible to prove convergence without it.

6. EMPIRICAL RESULTS

The algorithm with the step-size procedure and the stopping rule was applied to a recurring (weekly) problem in banking. The detailed formulation of the problem is not in the scope of this paper. We will give a brief description in general terms and, subsequently, we will state some observations regarding the performance of the algorithm, derived from hundreds of cases with real data.

The problem is concerned with the management of the bond portfolio of a bank. Its mathematical form is given by (1) and (2). The components of the vector $x = (x_1, \ldots, x_N)$ denote decisions at discrete points in time, with positive values signifying buying securities and negative values signifying selling securities. The profit function f in (1) is concave. The constants b_j in (2) are all equal to $-b$, where $b \geqslant 0$ denotes the initial quantity of securities on hand. Thus, the constraints in (2) state that the sum of all decisions up to each point in time cannot be smaller than the initial quantity of securities on hand, that is, the quantity on hand must always be nonnegative. These constraints are necessitated by the fact that banks cannot "short" their bond portfolios (i.e. borrow to sell a security at some point in time, and buy the same security later to return it to the original owner).

In addition to the regular weekly applications in the course of about 2 years, sensitivity tests were performed on the various user-specified parameters employed in the algorithm. The conclusions derived from hundreds of applications are summarized below :

1) An initial feasible point is required in step 0. It is not important to start with a good point; the algorithm makes giant steps in the first few iterations and soon attains a point that is nearly optimal.

2) The parameter b in the step-size procedure (step 2') can be set equal to any point in the (open) interval (0, 0.5). Larger b values generally mean more iterations for the step-size procedure, but fewer iterations for the algorithm itself. In the portfolio problem, these two opposite effects seem to cancel out one-another, so the speed of convergence appears independent of the b value used, for b in the interval [0.01, 0.49]. However, for b values close to 0.5, the convergence time would approach infinity as b approached 0.5, since then the set of acceptable points for the step size would approach a finite set.

3) In empirical applications of any iterative procedure, it is advisable to place an upper limit on the number of iterations. This is true, even if the procedure is finite and expected to converge within a few iterations, as appears to be the case with our step-size procedure, just in case something goes wrong. A limit of 25 on the number of iterations of this procedure appeared to be amply sufficient when applied to the portfolio problem.

4) The number of iterations to convergence of the algorithm depends, of course, on the values of the parameters δ and ϵ. The values of these parameters must be sufficiently small for the solution at termination to be satisfactorily close to a K-T point. However, they cannot be set equal to zero, since the algorithm would then terminate only if a K-T point has been attained, generally not in a finite number of iterations. In order to find appropriate values, it is necessary to perform a large number of sensitivity tests, parametrically reducing the values δ and ϵ, until further reductions do not substantially increase the value of the objective function at the terminal solution. This will depend on the structure of the particular problem under consideration. If obtaining better solutions becomes costly, the marginal benefit from a better solution must be balanced judgementally against the marginal cost. For the portfolio problem, $\delta = 0.0001$ and $\epsilon = 0.001$ appeared sufficiently good, although the additional cost of obtaining better solutions was negligible. With these specifications, the algorithm usually converged within 15 iterations, and the terminal objective function value appeared to be within 0.1 % of the optimal solution value (letting the algorithm run for more than 100 iterations would produce an objective function value smaller than 100.1 % of the value obtained with the stopping rule).

7. CONCLUSIONS AND FUTURE RESEARCH

The main points of this paper can be summarized as follows :

1) A direction vector that approximates the vector of Lagrange multipliers was used. This direction vector is not necessarily nonnegative, but is convergent to the vector of Lagrange multipliers in the limit.

2) The direction vector was bent in the one-dimensional subproblem every time a constraint became active, in such a way that movement ceased along components violating feasibility, but continued along the remaining components.

3) The one-dimensional subproblem of the algorithm was replaced by one that can be solved finitely.

4) A stopping rule that guarantees finite convergence to a point arbitrarily close to a K-T point was used.

The techniques used in this paper may be applicable in other situations.

The concept of using the Lagrangean vector as direction vector might prove useful in other cases where such a vector can be computed. McCormick [8] used the gradient as direction vector. It so happens that for his problem the Lagrangean vector is the gradient itself.

The concept of bending the direction vector in order to avoid infeasibility might be useful in modifying some feasible direction methods that are subject to the zigzagging phenomenon. Such a modification might also lead to faster convergence, since the existing option, that is, stopping movement in the one-dimensional subproblem when a direction component violates feasibility would also be available.

The modification of the one-dimensional subproblem is simple and could be applied to any algorithm that proceeds by selecting a direction and then optimizing along that direction. This modification is not dependent on concavity. It would be computationally of interest to apply a similar modification to some other method and then prove convergence.

The concept used in designing the finite stopping rule seeks to satisfy the K-T conditions and is not dependent on concavity. Loosely speaking, if an algorithm generates a solution sequence with a least one limit point and with every such point satisfying the K–T conditions, then, as a limit point is approached, the K–T conditions will come closer to being satisfied, provided that certain continuity conditions hold. This concept might be useful in designing stopping rules for other algorithms in nonlinear programming.

AKNOWLEDGMENT

This paper is an outgrowth of a Ph.D. dissertation [5], submitted to Yale University in 1973. I am thankful to the members of the dissertation committee, Professors Robert Mifflin (chairman), Martin Shubik and Harvey M. Wagner, for their guidance and assistance.

REFERENCES

1. CAZALET E. G. : "A tutorial development of decomposition for bond portfolio management", Stanford Research Institute Memorandum No. 22, Project 7799, 1972.
2. DANIEL J. W. : "Convergent step sizes for gradient-like feasible direction algorithms for constrained optimization", in J. B. Rosen, et. al. (eds.), *Nonlinear Programming,* Academic Press, New York, 1970, 245-274.
3. FRANK M. and P. WOLFE, "An algorithm for quadratic programming", *Naval Logistics Quarterly,* III (1956), 95-110.
4. GOLDSTEIN A. A. : *Constructive real analysis,* Harper, New York, 1967.
5. MAVRIDES L. P. : "Nonlinear programming under triangular constraint substructure, with applications to sequential decision processes", Ph.D. dissertation, Yale University, 1973.
6. MAVRIDES L. P. : "Optimization under nonnegativity constraints", Numerische Mathematik, 28 (1977), 287-293.
7. MAVRIDES L. P. : "An implementable F-W algorithm applied to a banking problem", INFOR, 16 (1978) 2.
8. McCORMICK G. P. : "Anti-zig-zagging by bending", Management Science, 15 (1969), 315-320.
9. ZANGWILL W. I. : *Nonlinear programming : a unified approach,* Prentice-Hall, Englewood Cliffs, N. J., 1969.